

QP Code :15387

(3 Hours)

[Total Marks : 100

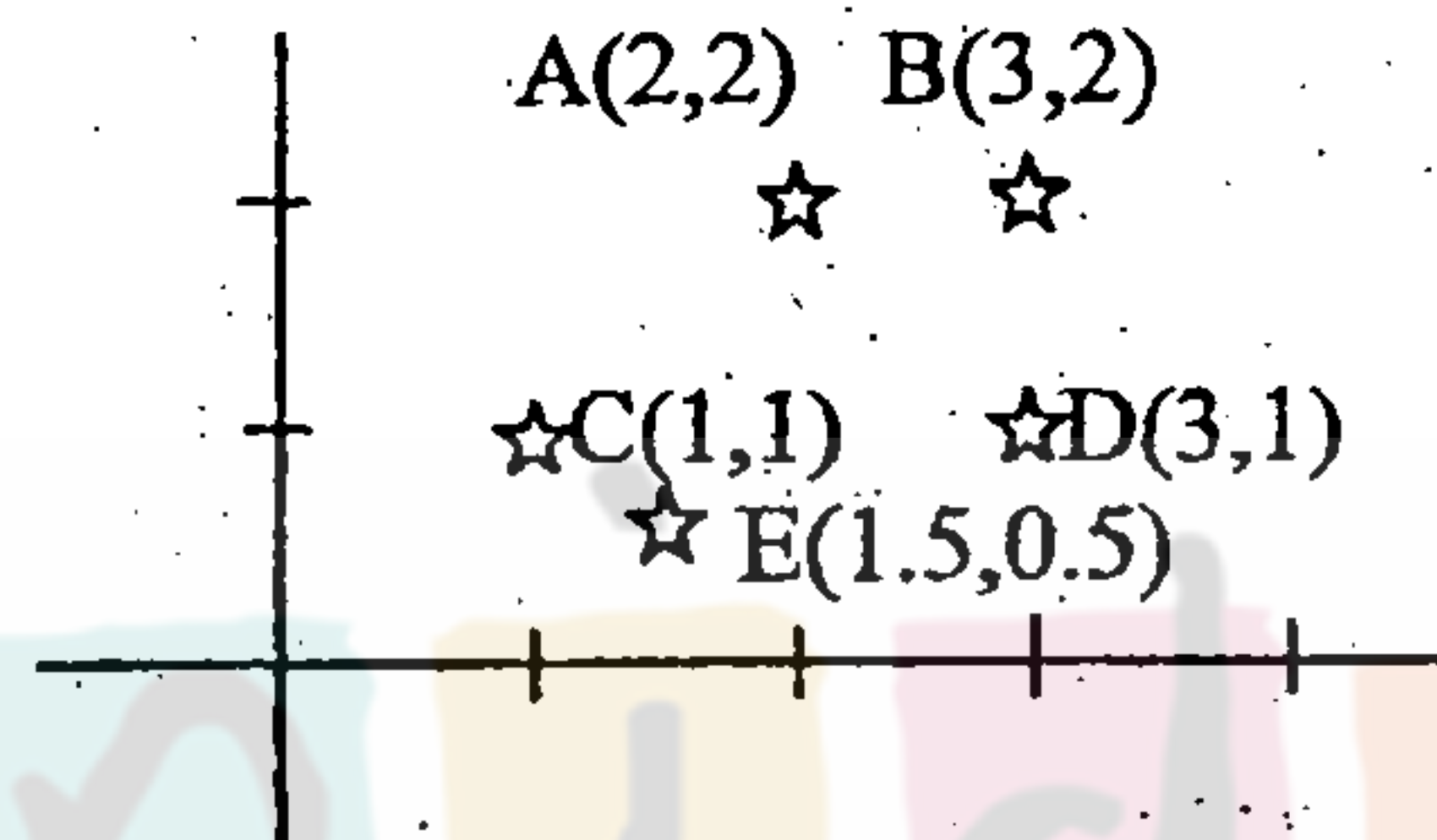
- N.B. :** (1) Question No. 1 is **compulsory**.
 (2) Attempt total **five** questions.
 (3) **Figures** to the **right** indicate **full** marks.

1. (a) What is Web Structure Mining? What are the techniques used for it ? 5
 (b) Compare OLTP & OLAP. 5
 (c) Consider the transaction database in the Appendix B. Use Apriori Algorithm with minimum support of 30% find all Frequent Item-sets. 5
 (d) What is multidimensional association rule ? 5
2. (a) A manufacturing Company has a huge sales network. To control the sales, it is divided in the regions. Each region has multiple zones. Each zone has different cities. Each sales person is allocated different cities. The object is to track Sales figure at different granularity levels of Region. Also to count number of products sold. Products are categorized as High end and low end products. Develop BI application, taking into consideration of above granularity levels for region, Sales person, Product and the Quarterly, yearly and monthly sales. Also it should predict Zone wise, product-wise sales for subsequent quarters.
 (i) Identify facts and dimensions and hence draw Information Package Diagram. 5
 (ii) Design suitable DWH schema. 5
 (iii) Identify suitable DM algorithm for predicting the sales. 5
 (iv) Give justification for all the decisions you have taken for the design. 5
3. (a) Use Data set in Appendix A. Create Adjacency Matrix. Use Single Link OR Complete Link Algorithm to Cluster given data set by drawing Dendogram. 10
 (b) With neat diagram explaining the process of KDD giving emphasis on selection and pre-processing phase. 10
4. (a) Use Data set in Appendix A. Use K-means algorithm to create two clusters. 10
 (b) What are the features required in Mining algorithm to Cluster Stream Data ? 10
 Explain any Clustering algorithm used for Stream Data.
5. (a) Explain Data Integration and Transformation w.r.t Data Warehouse. 10
 (b) Explain BIRCH algorithm with example. 10

[TURN OVER

6. (a) What is Concept Hierarchy ? How Concept Hierarchy is generated for Numerical and categorical data ? 10
- (b) Using table in Appendix C, Create classification model using decision-tree and hence classify following tuple. Very High, Old ? 10
7. Answer any **two** questions :- 20
- (a) What is text mining ? Explain different approaches of text mining.
- (b) What is CLICK-STREAM Mining ?
- (c) What are the applications of Web usage mining ? What is Web Log ? Give typical structure of web log ?

Appendix A.



Appendix B.

| TID | Items |
|-----|------------|
| 01 | A, B, D |
| 02 | B, C, D |
| 03 | A, B |
| 04 | B, D |
| 05 | A, B, C, D |

Appendix C

| No. | Income | Age | Own House |
|-----|-----------|--------|-----------|
| 1. | Very High | Young | Yes |
| 2. | High | Medium | Yes |
| 3. | Low | Young | Rented |
| 4. | High | Medium | Yes |
| 5. | Very High | Medium | Yes |
| 6. | Medium | Young | Yes |
| 7. | High | Old | Yes |
| 8. | Medium | Medium | Rented |
| 9. | Low | Medium | Rented |
| 10. | Low | Old | Rented |
| 11. | High | Young | Yes |